第27卷 第1期 工 程 数 学 学 报 Vol. 27 No. 1
2010年02月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Feb. 2010

文章编号:1005-3085(2010)01-0037-10

基于一般包含度的广义变精度粗糙集*

张红英[1]， 董鸣皋[2]

(1- 西安交通大学理学院，西安 710049; 2- 西安石油大学经管学院，西安 710065)

摘 要: 本文将变精度粗糙集模型推广到基于一般包含度的广义变精度粗糙集，利用包含度的性质讨论广义变精度粗糙集的性质。给出广义齐次包含度的定义，验证了常用的概率型包含度等均为广义齐次包含度；给出基于该广义变精度粗糙集的知识约简的方法。不同的包含度代表不同的决策语义，因此借助一般包含度研究变精度粗糙集，能够使尽可能多的有用信息被提取、挖掘，克服了基本粗糙集模型中由于要求绝对精确的包含关系而使大量有用信息丢失，进一步推广了粗糙集模型，拓宽了粗糙集在数据挖掘、知识发现、模式识别及决策分析等领域中的应用。

关键词: 变精度粗糙集；包含度；知识约简；协调集

分类号: AMS(2000) 03E02; 68T30 **中图分类号**: TP18 **文献标识码**: A

1 引言

知识发现是人工智能的核心问题之一，它是从信息系统中识别正确、新颖、有潜在应用价值并最终可为人们所理解的模式的方法。粗糙集理论[1] 是Pawlak教授于1982年提出的一种研究不完整、不确定知识和数据的表达、学习、归纳的理论方法，提供了知识发现的一种数学方法，实现了知识约简这一知识发现的重要课题。

一般信息系统的知识约简大多是在Pawlak粗糙集模型下进行的[1,2]。Pawlak粗糙集模型的一个局限性是它所处理的分类必须是完全正确的或肯定的，因而它的分类是精确的，亦即只考虑完全“包含”与“不包含”，而没有某种程度上的“包含”与“属于”。Pawlak粗糙集模型的另一个局限性是它所处理的对象是已知的，且从模型中得到的结论仅适用于这些对象。但在实际应用中，往往需要把从小规模对象集中得到的结论应用于大规模对象集上去。Pawlak粗糙集模型的这些局限性限制了它的应用。近年来，许多学者把这一模型推广至变精度粗糙集模型(VPRS)[3−6]。Ziarko[3] 于1993年利用一个阈值，规定对象所在的等价类在某种程度上包含于集合X中时，就认为这个对象属于X。这一推广在应用上是非常重要的，因为在实际问题中绝对的包含有时是苛刻的也是不必要的。基于变精度粗糙集理论，文献[3-6]给出并研究了不协调信息系统的β下近似约简。β下近似约简保持有决策的对象总数不变，但所产生的决策规则与原信息系统产生的规则有可能冲突。

包含度理论[2,7,8] 是对经典包含关系、序关系及蕴含关系等概念的定量化处理，因此结合包含度理论研究变精度粗糙集理论是一个重要的途径。包含度为衡量偏序集中任意两个元素之间大小程度的直观的测度。很多学者利用公理化方法[7] 和构造性方法[8-10] 研究包含度。文献[11]指出粗糙集的上下近似定义、近似精度和粗糙隶属度都归结为包含度。文献[12]提出基

收稿日期: 2008-05-27. **作者简介**: 张红英(1979年1月生)，女，博士，讲师. 研究方向：模糊集、粗糙集与人工智能.
***基金项目**: 国家自然科学基金(60673096; 60773174; 60703117)；陕西省教育厅社科基金(07JK101).

于概率型包含度的变精度粗糙集模型上的β上、下分布约简。同时给出了β上分布约简与β下分布约简的判定定理和相应的可辨识属性矩阵，从而得到了变精度粗糙集模型上知识约简的新方法。文献[2,8,13]利用模糊集的包含度定义带模糊决策的模糊信息系统的上、下近似约简，决策约简，最大决策约简等。孙士保等人[14]将变精度粗糙集中的等价关系推广到一般二元关系，提出基于后继邻域的广义变精度粗糙集模型。Hong[15]等人利用变精度粗糙集模型在具有噪声和不确定性的海量数据中挖掘$\beta-$确定规则和$\beta-$近似规则。

然而，到目前为止，已有的变精度粗糙集模型的讨论中，主要是借助概率型包含度衡量粗糙集中的包含关系，为此，本文讨论基于一般包含度研究变精度粗糙集的性质及基于该种变精度粗糙集模型的知识约简。首先引入一般包含度上的变精度粗糙集，借助包含度的性质讨论广义变精度粗糙集的性质，建立两者之间的关系，不同的包含度生成具有不同性质的变精度粗糙集。引入广义齐次包含度，讨论基于广义变精度粗糙集的上下分布约简，近似约简，分配约简，讨论它们之间的关系并给出约简定理。

2 基本知识

设(U,A)为一个信息系统，其中$U=\{x_1,x_2,\cdots,x_n\}$为非空有限对象集，$\mathscr{P}(U)$为U上的全体经典子集组成的集合，$\mathscr{F}(U)$为U上的全体模糊集集合。$A=\{a_1,a_2,\cdots,a_m\}$为非空有限属性集，而$|A|$为集合A的势。

定义1[2,12] 设(U,A,D)为目标信息系统，对于$B\subseteq A$，R_B与R_D分别为U上由B与D导出的等价关系，它们分别产生的U上的分划为

$$U/R_B=\{[x]_B:x\in U\}\quad 与\quad U/R_D=\{[x]_D:x\in U\}=\{D_1,D_2,\cdots,D_r\},$$

其中

$$[x]_B=\{y\in U:(x,y)\in R_B\},\quad [x]_D=\{y\in U:(x,y)\in R_D\},$$

分别是x关于B与D的等价类。对于任意$X\subseteq U$，$\beta\in(0.5,1]$，记

$$\underline{R_{BP}}^{\beta}(X)=\left\{x\in U:\frac{|[x]_B\cap X|}{|[x]_B|}\geq\beta\right\}=\bigcup\left\{[x]_B:\frac{|[x]_B\cap X|}{|[x]_B|}\geq\beta\right\},$$

$$\overline{R_{BP}}^{\beta}(X)=\left\{x\in U:\frac{|[x]_B\cap X|}{|[x]_B|}>1-\beta\right\}=\bigcup\left\{[x]_B:\frac{|[x]_B\cap X|}{|[x]_B|}>1-\beta\right\}.$$

$\underline{R_{BP}}^{\beta}(X)$与$\overline{R_{BP}}^{\beta}(X)$分别称为$X$关于$B$的$\beta$下近似和$\beta$上近似。称$(\underline{R_{BP}}^{\beta}(X),\overline{R_{BP}}^{\beta}(X))$为变精度粗糙集。

当$\beta=1$时，

$$\underline{R_{BP}}^{\beta}(X)=\underline{R_B}(X),\quad \overline{R_{BP}}^{\beta}(X)=\overline{R_B}(X).$$

因此，变精度粗糙集是Pawlak粗糙集的推广。当$D=\{d\}$时，在变精度粗糙集模型中，若$x\in\underline{R_{AP}}^{\beta}(D_j)$，其中$D_j=\{x\in U:d(x)=j\}$时，与之对应有不确定性命题规则

$$\bigwedge_{l=1}^{m}(a_l,a_l(x))\to d=j.$$

容易验证文献[12]中定义的变精度上、下近似算子$\underline{R_{BP}}^{\beta}$与$\overline{R_{BP}}^{\beta}$具有下述性质。

1) $\underline{R_{BP}}^{\beta}(X) =\sim \overline{R_{BP}}^{\beta}(\sim X)$，对任意的 $X \subseteq U$；

2) $\underline{R_{BP}}^{\beta}(D_i) \cap \underline{R_{BP}}^{\beta}(D_j) = \emptyset$, $i \neq j$；

3) $\bigcup\limits_{j=1}^{r} \underline{R_{BP}}^{\beta}(D_j) \subseteq \bigcup\limits_{j=1}^{r} \overline{R_{BP}}^{\beta}(D_j)$；

4) $\overline{R_{BP}}^{\beta}(D_i) \cap \overline{R_{BP}}^{\beta}(D_j) = \emptyset$, $i \neq j$，一般不成立。其中 $\sim X$ 为 X 的补集。

显然，基于变精度粗糙集理论，并不是每一个对象都有决策，只有当其所在等价类在某个决策类中的包含度不小于阈值时，它才能产生决策规则。

定义 2 设 (U, A, D) 是目标信息系统，若 $R_A \subseteq R_D$，则称目标信息系统是协调的，否则称目标信息系统是不协调的。

定义 3[7] 令 $\mathscr{F}(U)$ 上的包含度 I 是其上的映射：$\mathscr{F}(U) \times \mathscr{F}(U) \longrightarrow [0,1]$。对于任意的 $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathscr{F}(U)$，$I$ 满足以下公理。

公理 1 $I(\mathcal{X},\mathcal{Y}) = 1 \Leftrightarrow \mathcal{X} \subseteq \mathcal{Y}$；

公理 2 $I(\mathcal{X},\mathcal{Y}) = 0$，如果 $\mathcal{X} = U, \mathcal{Y} = \emptyset$；

公理 3 $\mathcal{X} \subseteq \mathcal{Y} \Rightarrow I(\mathcal{Y},\mathcal{Z}) \leq I(\mathcal{X},\mathcal{Z})$，即包含度对第一个变量是单调递减的；

公理 4 $\mathcal{X} \subseteq \mathcal{Y} \Rightarrow I(\mathcal{Z},\mathcal{X}) \leq I(\mathcal{Z},\mathcal{Y})$，即包含度对第二个变量是单调递增的；

公理 5 $I(\mathcal{X},\mathcal{Y}) = I(\sim \mathcal{X}, \sim \mathcal{Y})$；

公理 6 $I(\mathcal{X} \cup \mathcal{Y},\mathcal{Z}) = I(\mathcal{X},\mathcal{Z}) \wedge I(\mathcal{Y},\mathcal{Z})$；

公理 7 $I(\mathcal{X},\mathcal{Y} \cap \mathcal{Z}) = I(\mathcal{X},\mathcal{Y}) \wedge I(\mathcal{X},\mathcal{Z})$。

当然一般的包含度不一定全部满足上述公理，而只满足一部分公理。例如概率型包含度

$$I_1(\mathcal{X},\mathcal{Y}) = \begin{cases} \frac{|\mathcal{X}\cap\mathcal{Y}|}{|\mathcal{X}|}, & \mathcal{X} \neq \emptyset, \\ 1, & \mathcal{X} = \emptyset. \end{cases}$$

满足公理 1, 2, 4，但是不满足其它公理。因此文 [2] 给出了一般包含度的概念，若 I 满足

1) $0 \leq I(\mathcal{X},\mathcal{Y}) \leq 1$；　2) 若 $\mathcal{X} \subseteq \mathcal{Y}$，有 $I(\mathcal{X},\mathcal{Y}) = 1$；

3) 若 $\mathcal{X} \subseteq \mathcal{Y} \subseteq \mathcal{Z}$，有 $I(\mathcal{Z},\mathcal{X}) \leq I(\mathcal{Z},\mathcal{Y})$。

定义 1 借助概率型包含度 $I_1(\mathcal{X},\mathcal{Y})$ 定义了变精度近似算子，其中 $\mathcal{X}, \mathcal{Y} \in \mathscr{P}(U)$。

3 基于一般包含度的广义变精度粗糙集

在这一节，我们将变精度粗糙集中的概率型包含度推广为一般包含度，讨论广义变精度粗糙集的性质和包含度性质之间的关系，进一步讨论基于该种广义变精度粗糙集模型的知识约简。

基于 $\mathscr{P}(U)$ 上一般包含度 I 的广义变精度上、下近似算子定义如下，其中 $\beta \in (0.5, 1]$。

$$\underline{R_B}^{\beta}(X) = \{x \in U : I([x]_B, X) \geq \beta\} = \bigcup \{[x]_B : I([x]_B, X) \geq \beta\},$$

$$\overline{R_B}^{\beta}(X) = \{x \in U : I([x]_B, X) > 1 - \beta\} = \bigcup \{[x]_B : I([x]_B, X) > 1 - \beta\}.$$

容易验证广义变精度近似算子 $\underline{R_B}^{\beta}$ 与 $\overline{R_B}^{\beta}$ 满足下述性质。

性质 1 设 (U, A, D) 为目标信息系统，对于 $B \subseteq A$，U/R_B 与 U/R_D 分别为由 U 上的等价关系 R_B 与 R_D 导出的 U 上的分划，即

$$U/R_B = \{[x]_B : x \in U\}, \quad U/R_D = \{[x]_D : x \in U\} = \{D_1, D_2, \cdots, D_r\},$$

则广义变精度近似算子 $\underline{R_B}^\beta$ 和 $\overline{R_B}^\beta$ 具有如下性质。

1) $\underline{R_B}^\beta(X) = \sim \overline{R_B}^\beta(\sim X)$，对任意的 $X \subseteq U$ 当且仅当 $I([x]_B, X) + I([x]_B, \sim X) \geq 1$；

2) $\underline{R_B}^\beta(X) \subseteq \overline{R_B}^\beta(X)$；

3) $\underline{R_B}^1(X) = \underline{R_B}(X)$, $\overline{R_B}^1(X) = \overline{R_B}(X)$ 当且仅当 $I(X,Y) \neq 0$,$X \cap Y \neq \emptyset$；

4) 若 $I(\cdot,\cdot)$ 关于第二个变量是单调递增的，对于 $X_1 \subseteq X_2$，则有

$$\underline{R_B}^\beta(X_1) \subseteq \underline{R_B}^\beta(X_2), \quad \overline{R_B}^\beta(X_1) \subseteq \overline{R_B}^\beta(X_2);$$

5) 若 $I(\cdot,\cdot)$ 关于第一个变量是单调递减的，对于 $B_1 \subseteq B_2$，则有

$$\underline{R_{B_1}}^\beta(X) \subseteq \underline{R_{B_2}}^\beta(X), \quad \overline{R_{B_1}}^\beta(X) \subseteq \overline{R_{B_2}}^\beta(X);$$

6) 若 $I([x]_B, X) + I([x]_B, \sim X) = 1$，且 $I(\cdot,\cdot)$ 关于第二个变量是单调递增的，则有

$$\underline{R_B}^\beta(D_i) \cap \overline{R_B}^\beta(D_j) = \emptyset, \quad i \neq j;$$

7) $\bigcup\limits_{j=1}^{r} \underline{R_B}^\beta(D_j) \subseteq \bigcup\limits_{j=1}^{r} \overline{R_B}^\beta(D_j)$；

8) $\overline{R_B}^\beta(D_i) \cap \overline{R_B}^\beta(D_j) = \emptyset$, $i \neq j$，一般不成立。

证明 只证性质 (6)，其它性质易得。

若对于任意的 $[x]_B \subseteq \underline{R_B}^\beta(D_i)$，即 $I([x]_B, D_i) \geq \beta > 0.5$。由于

$$I([x]_B, D_i) + I([x]_B, \sim D_i) = 1,$$

故 $I([x]_B, D_i) \leq 1 - \beta$。因为 $i \neq j$，因此有 $D_i \cap D_j = \emptyset$，故 $D_j \subseteq \sim D_i$，结合 I 的单调性，有 $I([x]_B, D_j) \leq I([x]_B, \sim D_i) \leq 1 - \beta < 0.5$，即 $I([x]_B, D_j) \geq \beta$ 不成立，综上可知 $\underline{R_B}^\beta(D_i) \cap \overline{R_B}^\beta(D_j) = \emptyset$, $i \neq j$。

注1 概率型包含度 $I_1(\mathcal{X},\mathcal{Y})$ 关于第二个变量单调递增，同时满足 $I_1([x]_B, X) + I_1([x]_B, \sim X) = 1$；包含度

$$I_2(\mathcal{X},\mathcal{Y}) = \begin{cases} \frac{|\mathcal{Y}|}{|\mathcal{X} \cup \mathcal{Y}|}, & \mathcal{X},\mathcal{Y} \neq \emptyset, \\ 1, & \mathcal{X} = \emptyset. \end{cases}$$

关于第一个变量是单调递减的，关于第二个变量是单调递增的。

下面我们介绍基于一般包含度的广义变精度粗糙集的知识约简的基本概念及它们之间的关系。

定义4 设 (U,A,D) 为目标信息系统，对于 $B \subseteq A$。记

$$\sigma_B^\beta = \frac{\sum\left\{\left|\underline{R_B}^\beta(D_j)\right| : \ j \leq r\right\}}{|U|}, \qquad \lambda_B^\beta = \frac{\sum\left\{\left|\overline{R_B}^\beta(D_j)\right| : \ j \leq r\right\}}{|U|},$$

$$L_B^\beta = \left(\underline{R_B}^\beta(D_1), \cdots, \underline{R_B}^\beta(D_r)\right), \qquad H_B^\beta = \left(\overline{R_B}^\beta(D_1), \cdots, \overline{R_B}^\beta(D_r)\right),$$

$$\mu_B(x) = \left(I([x]_B, D_1), \cdots, I([x]_B, D_r)\right), \qquad \delta_B(x) = \left\{D_j : I([x]_B, D_j) > 0\right\},$$

$$\gamma_B(x) = \left\{D_k : I([x]_B, D_k) = \max_{j \leq r}\{I([x]_B, D_j)\}\right\},$$

1)　若 $\sigma_B^\beta = \sigma_A^\beta$，则称 B 是目标信息系统 (U, A, D) 的 β 下近似协调集。若 B 是 β 下近似协调集，但 B 的任何真子集不是 β 下近似协调集，则称 B 是目标信息系统 (U, A, D) 的 β 下近似约简。

2)　类似地，定义目标信息系统 (U, A, D) 的 β 上近似协调集和 β 上近似约简；β 下分布协调集和 β 下分布约简；β 上分布协调集和 β 上分布约简；β 分布协调集和 β 分布约简；β 最大分布协调集和 β 最大分布约简；β 分配协调集和 β 分配约简。

注 2　定义 4 中的前六种约简方法是文献 [2,12] 中介绍的方法的直接推广，即将原来的概率型包含度替换为一般包含度。

β 上 (下) 分布协调集是保持每个决策类的 β 上 (下) 近似不变的属性集，它与 A 产生相容的命题规则，即在原系统和约简系统中，由同一对象所产生的命题规则的决策部分相同；而 β 下近似协调集保持决策类的下近似中的对象总数不变，由它产生的命题规则与由 A 产生的命题规则可能冲突，但支持这些命题规则的对象个数不变；分布协调集是保持对象在每个决策类的隶属程度不变的属性集，而最大分布协调集保持每个对象的最大分布决策类不变；对于协调目标信息系统，上述约简是一致的。

根据定义 4 立即可得定理 1。

定理 1　设 (U, A, D) 为目标信息系统，则 β 下分布协调集必为 β 下近似协调集，β 上分布协调集必为 β 上近似协调集。

注 3　定理 1 的逆命题不一定成立，文献 [2,12] 在概率型包含度定义的粗糙集上给出反例。

定理 2　设 (U, A, D) 为目标信息系统，则分布协调集必为 β 上、下分布协调集。

这是文献 [2] 中的定理 2.34.

为了进一步讨论上述约简之间的关系，我们引进广义齐次包含度的定义。

定义 5　令 I 为 $\mathscr{F}(U)$ 上的包含度。I 称为广义齐次包含度，若对于任意的非空集合 $X \in \mathscr{P}(U)$，存在 X_1，$X_2 \in \mathscr{P}(U)$，其中 $X = X_1 \cup X_2$，$X_1 \cap X_2 = \emptyset$，I 满足。

对于任意的集合 $\mathcal{Y} \in \mathscr{F}(U)$，若 $I(X_1, \mathcal{Y}) = a_1, I(X_2, \mathcal{Y}) = a_2$，则有 $I(X, \mathcal{Y}) \geq \min\{a_1, a_2\}$。

注 4　广义齐次包含度是文献 [16] 中定义 6.7 定义的齐次包含度的一种推广形式。

定理 3　以下 $\mathscr{F}(U)$ 上的包含度均为广义齐次包含度

$$I_1(\mathcal{X}, \mathcal{Y}) = \begin{cases} \frac{|\mathcal{X} \cap \mathcal{Y}|}{|\mathcal{X}|}, & \mathcal{X} \neq \emptyset, \\ 1, & \text{其它}. \end{cases} \qquad I_3(\mathcal{X}, \mathcal{Y}) = \bigwedge_{x \in U} (\mathcal{X}^c(x) \vee \mathcal{Y}(x)),$$

$$I_4(\mathcal{X}, \mathcal{Y}) = \bigvee_{x \in U} (\mathcal{X}(x) \wedge \mathcal{Y}(x)), \qquad I_5(\mathcal{X}, \mathcal{Y}) = \bigwedge_{x \in U} \{\mathcal{X}^c(x) + \mathcal{Y}(x) - \mathcal{X}^c(x)\mathcal{Y}(x)\},$$

$$I_6(\mathcal{X}, \mathcal{Y}) = \bigwedge_{x \in U} \{(\mathcal{X}^c(x) + \mathcal{Y}(x)) \wedge 1\}, \quad I_7(\mathcal{X}, \mathcal{Y}) = \bigwedge_{x \in U} \left\{\frac{\mathcal{Y}(x)}{\mathcal{X}(x)} \wedge 1\right\},$$

$$I_8(\mathcal{X}, \mathcal{Y}) = \frac{|\varphi(\mathcal{X}, \mathcal{Y})|}{|\text{supp}\mathcal{X}|},$$

其中

$$\text{supp}X = \{x \in U : \mathcal{X}(x) > 0\}, \quad \varphi(\mathcal{X}, \mathcal{Y}) = \{x \in \text{supp}\mathcal{X} : \mathcal{X}(x) \leq \mathcal{Y}(x)\}.$$

证明　只证 I_1 为广义齐次包含度，其它证明类似。对于任意的 X_1，$X_2 \in \mathscr{P}(U)$，其中 $X =$

$X_1 \cup X_2$，$X_1 \cap X_2 = \emptyset$，故

$$I(X_1,\mathcal{Y}) = \frac{\sum\limits_{x\in X_1} \mathcal{Y}(x)}{|X_1|}, \qquad I(X_2,\mathcal{Y}) = \frac{\sum\limits_{x\in X_2} \mathcal{Y}(x)}{|X_2|},$$

$$I(X_1 \cup X_2,\mathcal{Y}) = \frac{\sum\limits_{x\in X_1} \mathcal{Y}(x) + \sum\limits_{x\in X_2} \mathcal{Y}(x)}{|X_1| + |X_2|}.$$

不妨设

$$I(X_1,\mathcal{Y}) = \frac{\sum\limits_{x\in X_1} \mathcal{Y}(x)}{|X_1|} \le I(X_2,\mathcal{Y}) = \frac{\sum\limits_{x\in X_2} \mathcal{Y}(x)}{|X_2|},$$

因而

$$\begin{aligned} I(X_1 \cup X_2,\mathcal{Y}) - I(X_1,\mathcal{Y}) &= \frac{\sum\limits_{x\in X_1} \mathcal{Y}(x) + \sum\limits_{x\in X_2} \mathcal{Y}(x)}{|X_1| + |X_2|} - \frac{\sum\limits_{x\in X_1} \mathcal{Y}(x)}{|X_1|} \\ &= \frac{|X_1| \sum\limits_{x\in X_2} \mathcal{Y}(x) - |X_2| \sum\limits_{x\in X_1} \mathcal{Y}(x)}{|X_1|(|X_1| + |X_2|)} \ge 0, \end{aligned}$$

根据假设，因此有 $I(X,\mathcal{Y}) \ge \min\{I(X_1,\mathcal{Y}),\ I(X_2,\mathcal{Y})\}$。

注5 由于 I_3, I_5, I_6, I_7 对于第一变量具有单调递减性，因此

$$I_i(X,\mathcal{Y}) = \min\{I_i(X_1,\mathcal{Y}), I_i(X_2,\mathcal{Y})\}, \quad i = 3,5,6,7.$$

当然并不是所有的包含度都是广义齐次包含度。例如根据 $I_2(\mathcal{X},\mathcal{Y})$ 的单调性，有

$$I_2(X_1 \cup X_2,\mathcal{Y}) \le \min\{I_2(X_1,\mathcal{Y}), I_2(X_2,\mathcal{Y})\},$$

且令 $X_1 = \{x_1,x_2\}$，$X_2 = \{x_3,x_5,x_7\}$，$\mathcal{Y} = \{x_1,x_4,x_5,x_6\}$，得

$$I_2(X_1,\mathcal{Y}) = \frac{4}{5}, \quad I_2(X_2,\mathcal{Y}) = \frac{2}{3}, \quad I_2(X_1 \cup X_2,\mathcal{Y}) = \frac{4}{7},$$

显然 $I_2(X_1 \cup X_2,\mathcal{Y}) < \min\{I_2(X_1,\mathcal{Y}), I_2(X_2,\mathcal{Y})\}$，因此 I_2 不是广义齐次包含度。

定理4 设 (U,A,D) 为目标信息系统，$B \subseteq A$。定义

$$\beta_B = \min\{\max I([x]_B, D_j) : x \in U\}, \quad \beta_0 = \min\{\beta_A, \beta_B\},$$

如果 $\beta_0 > 0.5$，且 $\{D_j : I([x]_B, D_j) > 0.5\}$ 为单元素集，则有

1) 对于 $\beta \in [0.5,\beta_0]$，若 B 是一个 β 下分布协调集，则 B 是一个最大分布协调集；

2) 对于 $\beta \in [0.5,\beta_A]$，包含度 I 是一个广义齐次包含度，若 B 是一个最大分布协调集，则 B 是一个 β 下分布协调集；

3) 对于 $\beta \in [0.5,\beta_0]$，包含度 I 是一个广义齐次包含度，若 B 是一个最大分布约简，则 B 是一个 β 下分布约简。

证明 1) 类似于文献[12]中定理3.3中(1)的证明。

2) 设 B 是一个最大分布协调集，根据定义有 $\gamma_B(x) = \gamma_A(x)$。而且由于 $B \subseteq A$，显然 $\{[y]_A : [y]_A \subseteq [x]_B\}$ 形成 $[x]_B$ 的一个划分。

对于任意的$j \le r$，由于$\{D_j : I([x]_B, D_j) > 0.5\}$为单元素集，故有

$$x \in \underline{R_A}^{\beta}(D_j) \Rightarrow I([x]_A, D_j) \ge \beta > 0.5 \Rightarrow \gamma(A)(x) = \{D_j\} \Rightarrow \gamma(B)(x) = \{D_j\}.$$

对于任意的$[y]_A \in \{[y]_A : [y]_A \subseteq [x]_B\}$，有$\gamma_B(x) = \gamma_B(y)$，结合$\gamma_B(y) = \gamma_A(y)$，得到$\gamma_B(x) = \gamma_A(y) = \{D_j\}$，即$I([y]_A, D_j) \ge \beta_A > \beta$。由于$I$是广义齐次包含度，故有

$$I([x]_B, D_j) \ge \min I([y]_A, D_j) > \beta,$$

因此有$x \in \underline{R_B}^{\beta}(D_j)$，同样的方法可以证明$\underline{R_B}^{\beta}(D_j) \subseteq \underline{R_A}^{\beta}(D_j)$，故有$\underline{R_B}^{\beta}(D_j) = \underline{R_A}^{\beta}(D_j)$。

利用1)和2)，易得结论3)成立。

定理5 设(U, A, D)为目标信息系统，$B \subseteq A$。定义

$$\alpha_B(D_j) = \min\{I([x]_B, D_j) : x \in U, I([x]_B, D_j) > 0\}, \quad \alpha_0(D_j) = \min\{\alpha_A, \alpha_B\},$$

则下列结论成立：

1) 对于$\beta \in (1 - \alpha_0(D_j), 1]$，若$B$是一个分配协调集，则$B$是一个$\beta$上分布协调集；

2) 对于$\beta \in (1 - \alpha_B(D_j), 1]$，若$B$是一个$\beta$上分布协调集，则$B$是一个分配协调集；

3) 对于$\beta \in (1 - \alpha_0(D_j), 1]$，若$B$是一个分配约简当且仅当$B$是一个$\beta$上分布约简。

证明 1) 若B是一个分配协调集，根据定义，对任意的$x \in U$有$\delta_B(x) = \delta_A(x)$，因此

$$x \in \overline{R_B}^{\beta}(D_j) \Rightarrow I([x]_A, D_j) > 1 - \beta > 0 \Rightarrow D_j \in \delta_B(x) \Rightarrow D_j \in \delta_A(x)$$

$$\Rightarrow I([x]_A, D_j) > 0 \Rightarrow I([x]_A, D_j) \ge \alpha_0(D_j) > 1 - \beta \Rightarrow x \in \overline{R_A}^{\beta}(D_j).$$

反之，同样的方法可以证明$x \in \overline{R_A}^{\beta}(D_j) \Rightarrow x \in \overline{R_B}^{\beta}(D_j)$，故有$\overline{R_A}^{\beta}(D_j) = \overline{R_B}^{\beta}(D_j)$。即$B$是一个$\beta$上分布协调集。

2) 若B是一个β上分布协调集，$\beta \in (1 - \alpha_B(D_j), 1]$，对任意的$j \le r$，有$\overline{R_A}^{\beta}(D_j) = \overline{R_B}^{\beta}(D_j)$。对于任意的$x \in U$，有

$$D_j \in \delta_B(x) \Rightarrow I([x]_B, D_j) \ge \alpha_B(D_j) > 1 - \beta \Rightarrow x \in \overline{R_B}^{\beta}(D_j)$$

$$\Rightarrow x \in \overline{R_A}^{\beta}(D_j) > 1 - \beta > 0 \Rightarrow D_j \in \delta_A(x).$$

另一方面，根据$\delta_B(x)$的定义，显然有$\delta_A(x) \subseteq \delta_B(x)$。因此对任意的$x \in U$有$\delta_A(x) \subseteq \delta_B(x)$，即$B$是一个分配协调集。

3) 结合1)和2)易得对于$\beta \in (1 - \alpha_0(D_j), 1]$，若$B$是一个分配约简当且仅当则$B$是一个$\beta$上分布约简。

一般地，如果β不满足上面的条件，那么一个β上分布约简不一定是分配约简，分配约简也不一定是上分布约简，反例见文献[12]。

4 基于广义变精度粗糙集的知识约简方法

在这一部分，我们研究基于广义变精度粗糙集模型的上、下分布约简方法。给出各种协调集下的等价描述。

定理6 设(U, A, D)是目标信息系统，I是广义齐次包含度，$B \subseteq A$。记

$$M_B^{\beta}(x) = \{D_j : x \in \overline{R_B}^{\beta}(D_j)\}, \quad \forall x \in U, \qquad G_B^{\beta}(x) = \{D_j : x \in \underline{R_B}^{\beta}(D_j)\}, \quad \forall x \in U,$$

则

1) B 是 β 上分布协调集 $\Leftrightarrow M_B^\beta(x) = M_A^\beta(y)$，对任意的 $x \in U$。

2) B 是 β 下分布协调集 $\Leftrightarrow G_B^\beta(x) = G_A^\beta(y)$，对任意的 $x \in U$。

证明 1) 由于 $D_j \in M_A^\beta(x) \Leftrightarrow x \in \overline{R_A}^\beta(D_j)$，易得结论 1), 2) 类似于 1) 可证。

下面给出基于广义变精度粗糙集模型的知识约简方法。

定理 7 设 (U, A, D) 是目标信息系统，$B \subseteq A$，I 是广义齐次包含度。则

1) B 是 β 上分布协调集 $\Leftrightarrow$ 任意的 $x, y \in U$，满足 $M_A^\beta(x) \neq M_A^\beta(y)$ 时，$[x]_B \cap [y]_B = \emptyset$。

2) B 是 β 下分布协调集 $\Leftrightarrow$ 任意的 $x, y \in U$，满足 $G_A^\beta(x) \neq G_A^\beta(y)$ 时，$[x]_B \cap [y]_B = \emptyset$。

证明 由于 $B \subseteq A$，显然有 $\{[y]_A : [y]_A \subseteq [x]_B\}$ 形成 $[x]_B$ 的一个划分。

1) $\Rightarrow$ 如果 $[x]_B \cap [y]_B \neq \emptyset$，对任意的 $x, y \in U$，有 $[x]_B = [y]_B$，故 $M_B^\beta(x) = M_B^\beta(y)$。根据定理 6，$B$ 是 β 上分布协调集 $\Leftrightarrow M_B^\beta(x) = M_A^\beta(x)$, $M_B^\beta(y) = M_A^\beta(y)$，因此 $M_A^\beta(x) = M_A^\beta(y)$。

$\Leftarrow$ 对于任意的 $x \in U$，若 $[y]_A \subseteq [x]_B$，有 $[x]_B \cap [y]_B \neq \emptyset$，根据假设有 $M_A^\beta(x) = M_A^\beta(y)$。

对于任意的 $j \leq r$，若 $x \in \overline{R_B}^\beta(D_j)$，即 $[x]_B \subseteq \overline{R_B}^\beta(D_j)$。既然

$$[x]_B = \cup\big\{[y]_A : [y]_A \in \{[y]_A : [y]_A \subseteq [x]_B\}\big\},$$

故对于任意的 $[y_0]_A \in \{[y]_A : [y]_A \subseteq [x]_B\}$，有 $[y_0]_A \subseteq \overline{R_A}^\beta(D_j)$，即 $D_j \in M_A^\beta(y_0)$。因此根据假设有 $D_j \in M_A^\beta(x)$，结合定理 6，有 $x \in \overline{R_A}^\beta(D_j)$。

反之，假设若 $x \in \overline{R_A}^\beta(D_j)$，即有 $D_j \in M_A^\beta(x)$。故对任意的 $[y]_A \in \{[y]_A : [y]_A \subseteq [x]_B\}$，有 $D_j \in M_A^\beta(y)$，即 $I([y]_A, D_j) > 1 - \beta$。又因为包含度 I 是一个广义齐次包含度，因此有

$$I([x]_B, D_j) \geq \min\big\{I([y]_A, D_j) : [y]_A \in \{[y]_A : [y]_A \subseteq [x]_B\}\big\} > 1 - \beta.$$

因此有 $x \in \overline{R_B}^\beta(D_j)$。即 $\overline{R_B}^\beta(D_j) = \overline{R_A}^\beta(D_j)$, B 是 β 上分布协调集。

2) 类似于 1) 的证明可得。

定理 7 给出了判断属性子集是 β 上、下分布协调集的方法。由此我们可进一步得到相应的知识约简方法。先给出可辨识属性矩阵的概念。

定义 6 设 (U, A, D) 是目标信息系统，$U/R_A = \{C_1, \cdots, C_m\}$。记

$$D_1^{*\beta} = \{([x]_A, [y]_A) : M_A^\beta(x) \neq M_A^\beta(y)\},$$

$$D_2^{*\beta} = \{([x]_A, [y]_A) : G_A^\beta(x) \neq G_A^\beta(y)\},$$

对于 $l = 1, 2$，定义辨识矩阵如下

$$D_l^\beta(C_i, C_j) = \begin{cases} \{a_k \in A : a_k(C_i) \neq a_k(C_j)\}, & (C_i, C_j) \in D_l^{*\beta}, \\ A, & (C_i, C_j) \notin D_l^{*\beta}, \end{cases}$$

根据辨识矩阵，类似于文献 [12] 证明方法，可以得到如下约简定理。

定理 8 辨识矩阵 $D_l^\beta = (D_l^\beta(C_i, C_j), i, j \leq m)$, $(l = 1, 2)$ 满足如下性质。

1) 矩阵是对称的，即 $D_l^\beta(C_i, C_j) = D_l^\beta(C_j, C_i)$，对任意的 $i, j \leq m$；

2) 主对角线上的元素都是 A，即 $D_l^\beta(C_i, C_i) = A$，对任意的 $i \leq m$；

3) $D_l^\beta(C_i, C_j) \subseteq D_l^\beta(C_i, C_s) \subseteq D_l^\beta(C_s, C_j)$，对任意的 $i, s, j \leq m$。

证明类似于文献 [12] 中定理 4.3 的证明。同样可得。

定理 9 设 (U, A, D) 是目标信息系统，$B \subseteq A$。则

1) B是一个β上分布协调集$\Leftrightarrow$对于任意的$(C_i,C_j)\in D_1^{*\beta}$，$B\cap D_1^\beta(C_i,C_j)\neq\emptyset$；

2) B是一个β下分布协调集$\Leftrightarrow$对于任意的$(C_i,C_j)\in D_2^{*\beta}$，$B\cap D_2^\beta(C_i,C_j)\neq\emptyset$；

定义7 设(U,A,D)是目标信息系统，$D_l^\beta=(D_l^\beta(C_i,C_j),i,j\leq m)$，$(l=1,2)$分别是$\beta$上下分布辨识矩阵。分别称

$$\begin{aligned}M_l^\beta &= \bigwedge\Big\{\bigvee\{a_k:a_k\in D_l^\beta(C_i,C_j)\}:i,j\leq m\Big\}\\ &= \bigwedge\Big\{\bigvee\{a_k:a_k\in D_l^\beta(C_i,C_j)\}:(C_i,C_j)\in D_l^{*\beta}\Big\},\quad l=1,2,\end{aligned}$$

为β上、下分布辨识公式。

定理10 设(U,A,D)是目标信息系统，辨识公式M_l^β的极小析取范式为

$$M_l^\beta=\bigvee_{k=1}^{t}\Big(\bigwedge_{s=1}^{q_k}a_{i_s}\Big),\quad l=1,2,$$

记$B_{lk}=\{a_{i_s}:s=1,2,\cdots,q_k\}$，则$\{B_{lk}:k=1,2,\cdots,t,(l=1,2)\}$分别是所有$\beta$上、下分布约简形成的集合。

证明类似文献[12]中定理4.5。

定理10提供了求不协调信息系统两种知识约简的方法。

5 结论

本文利用包含度的公理化定义研究广义变精度粗糙集的性质，满足一定性质的包含度保证了广义变精度近似算子的良好性质。引入广义齐次包含度的定义，在此基础上了讨论了基于变精度粗糙集的七种知识约简方法及它们之间的关系，说明上、下分布知识约简的判定定理和可辨识属性矩阵适用于广义变精度粗糙集模型，从而进一步拓宽了不协调目标信息系统知识约简的研究方向。

参考文献：

[1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1982, 11: 341-356

[2] 张文修，梁怡，吴伟志. 信息系统与知识发现[M]. 北京：科学出版社，2003
Zhang W X, Liang Y, Wu W Z. Information Systems and Knowledge Discovery[M]. Beijing: Science Press, 2003

[3] Ziarko W. Variable precision rough set model[J]. Journal of Computer and System Sciences, 1993, 46(1): 39-59

[4] Kryszkiewicz M. Comparative studies of alternative type of knowledge reduction in inconsistent systems[J]. International Journal of Intelligent Systems, 2001, 16: 105-120

[5] Quafatou M. α-RST: a generalization of rough set theory[J]. Information Sciences, 2000, 124: 301-316

[6] Beynon M. Reducts with in the variable precision rough sets model: a further investigation[J]. European Journal of Operational Research, 2001, 134: 592-605

[7] Sinha D, Dougherty E R. Fuzzification of set inclusion: theory and applications[J]. Fuzzy Sets and Systems, 1993, 55: 15-42

[8] 张文修，梁怡，徐萍. 基于包含度的不确定推理[M]. 北京：科学出版社，2007
Zhang W X, Liang Y, Xu P. The Uncertainty Reasoning Based on Inclusion Measures[M]. Beijing: Science Press, 2007

[9] Zhang H Y, Su Y J. A ranking approach with inclusion measure in multiple-attribute interval-valued decision making[C]// An A et al (eds), Proceeding of Rough Sets, Fuzzy Sets, Data Mining and Granular Computing 2007, LNAI 4482, 2007: 411-418

[10] Zhang H Y, Zhang W X. Inclusion measure and similarity measure of intuitionistic and interval-valued fuzzy sets[C]// Proceedings of the 2007 International Conference on Intelligent Systems and Knowledge Engineering (ISKE2007), Atlantis Press, PP. 2007: 415-421

[11] 梁吉业，李德玉. 信息系统的不确定性和知识获取[M]. 北京：科学出版社，2004
Liang J Y, Li D Y. The Uncertainty of Information Systems and Knowledge Discovery[M]. Beijing: Science Press, 2004

[12] Mi J S, Wu W Z, Zhang W X. Approaches to knowledge reduction based on variable precision rough set model[J]. Information Sciences, 2004, 159: 255-272

[13] Zhang H Y, Zhang W X. Using hybrid monotonic inclusion measure for knowledge reduction in information systems with fuzzy decisions[J]. Chinese Journal of Fuzzy Systems and Mathematics, 2007, 6: 141-147

[14] 孙士保，普杰信，秦克云. 广义变精度粗糙集模型中近似算子研究[J]. 计算机科学，2007, 34: 194-197
Sun S B, Pu J X, Qin K Y. The study on approximation operators of generalized variable precision rough sets[J]. Computer Science, 2007, 34: 194-197

[15] Hong T P, Wang T T, Wang S L. Mining fuzzy β-certain and β-possible rules from quantitative data based on the variable precision rough-set model[J]. Expert Systems with Applications, 2007, 32: 223-232

[16] 张文修，仇国芳. 基于粗糙集的不确定决策[M]. 北京：清华大学出版社，2005
Zhang W X, Qiu G F. Uncertain Decision Making Based on Rough Sets[M]. Beijing: Tsinghua University Press, 2005

Generalized Variable Precision Rough Sets Based on Normal Inclusion Measures

ZHANG Hong-ying[1], DONG Ming-gao[2]

(1- Faculty of Science, Xi'an Jiaotong University, Xi'an 710049;
2- School of Economy and Management, Xi'an Shiyou University, Xi'an 710065)

Abstract: This paper proposes approaches to the knowledge reduction based on generalized variable precision rough sets. The properties of generalized variable precision rough approximation operators are discussed based on the axiomatic definition of inclusion measure. A generalized homogeneous inclusion measure is introduced to study knowledge reduction based on a generalized variable precision rough set model. Since different inclusion measures mean various contexts, the generalized variable precision rough sets overcome the drawbacks which induce the loss of useful information in the classical rough set model and extend the usefulness and effectiveness of the rough set approach in data mining, knowledge discovery, pattern recognition, decision analysis and so on.

Keywords: variable precision rough sets; inclusion measure; knowledge reduction; consistent set

Received: 27 May 2008. **Accepted:** 10 Apr 2009.

Foundation item: The National Natural Science Foundation of China (60673096; 60703117; 60773174); the Social Science Foundation of the Education Department of Shaanxi Province of China (07JK101).